Co-inventor Networks and Knowledge Production in Specialized and Diversified Cities

Frank van der Wouden* **& David L. Rigby**
Department of Geography
University of California, Los Angeles

1255 Bunche Hall
Los Angeles, California
United States of America

*fvanderwouden@ucla.edu

PUBLISHED IN "PAPERS IN REGIONAL SCIENCE" 2019, v. 98 p. 1833-1853

Abstract

Research on the U.S. urban system has shown that metropolitan regions with more local and non-local network ties outperform cities where economic agents are isolated. Yet, little attention is given to the character of the local knowledge base and how that influences network structure. We show that co-inventor networks differ between cities that produce specialized and diversified knowledge. Models of tie-formation show inventors in specialized cities value spatial proximity less and cognitive proximity more than inventors in diversified cities as they partner with non-local inventors. These findings suggest that the influence of social networks on knowledge production is conditioned by the architecture of the local knowledge base.

Keywords: collaboration, knowledge production, patent, co-inventor, network, specialization

JEL: O33, R11, R12, R15, D85

1. Introduction

Knowledge production is increasingly imagined as an interactive task through which economic agents recombine existing ideas in novel ways (Arthur, 1999; Kauffman, 1993; Wuchty et al., 2007). Thus, the pace of invention relies upon access to multiple subsets of knowledge along with the capacity to translate those knowledge stocks into new technologies (Cohen & Levinthal, 1990). For economic geographers, these constraints on invention have historically focused attention on industrial districts or clusters within which actors generate economies from the reduced cost of interaction and from spillovers that are bounded by co-location (Jacobs, 1969; Marshall, 1920). Jaffe et al. (1993) and Audretsch and Feldman (1996) provide supporting empirical results. For economic sociologists, competitive advantage is "located" in the structure of social connections that economic agents cultivate (Powell, 1990; Uzzi, 1996). In this sense, social proximity is seen as independent of spatial proximity and perhaps more important in regulating the fortunes of firms and the flows of knowledge between them (see also Agrawal et al., 2006; Breschi & Lissoni, 2009)

Investigation of the geography of knowledge production problematizes the relationship between spatial, social and other forms of proximity, illustrating the conditions under which proximity is advantageous, but also when it becomes a liability (Boschma, 2005; Grabher, 1993). A good deal of this work contests the separation of spatial and social proximity, seeking to understand how co-location affects the structure of social ties (Broekel & Boschma, 2012; Gertler, 2003; Healy & Morgan, 2012). At the same time, the primacy of the local in the formation of social networks is questioned by Bathelt et al. (2004) and Amin and Cohendet (2004) who suggest that spatial embeddedness is less and less important to the inter-organizational linkages that enhance firm and regional performance.

In recent work on geographical variations in knowledge production, Fleming et al. (2007) and Lobo and Strumsky (2008) explore how the social networks that link inventors influence the pace of invention, independent of place-specific characteristics including agglomeration. Whittington et al. (2009) push this analysis further, examining the interaction between social and spatial relationships that influence innovation in knowledge-intensive industries. In this respect they add empirical detail to the earlier claims of Gordon and McCann (2000). Breschi and Lenzi (2016) update this work. Allied research by Cantner and Graf (2004) investigate some of these same themes while developing the connection between the technological specialization of regions and the structure of cooperation networks.

These ideas are extended in this paper that explores how co-inventor networks influence knowledge production in U.S. cities after controlling for a number of location-based covariates. We provide broad evidence that social networks and localized processes of agglomeration are positively related to knowledge production within cities, confirming earlier findings. At the scale of individual metropolitan areas, the advantages of spatial proximity and social proximity are shown to be substitutes for one another. We add value to existing research by illustrating how the structure of networks linking co-inventors within and between U.S. metropolitan areas depend on the nature of knowledge produced in different places (see also Cantner et al., 2010). Our results show that specialized cities, identified by high levels of cognitive proximity, develop significantly denser and more robust co-inventor networks than those found in diversified cities, and that such connectedness enhances knowledge production. In core models, explicit controls are employed to dampen the spatial dependence in U.S. metropolitan patent counts, extending previous work in this field. Finally, we report that the pipelines connecting inventors between cities vary with the level of metropolitan technological specialization: on average inventors in specialized cities are less impacted by geographical distance and more impacted by cognitive distance in their search for knowledge production partners in other urban areas.

The paper is organized in four following sections. Section 2 provides a brief review of the literature that motivates our research. In Section 3 we explore the operationalization of the core theoretical concepts and we discuss the sources of the data employed in our empirical analysis. The results from that analysis occupy Section 4 of the paper, and we offer a number of concluding remarks in Section 5.

2. Literature Review

Across the market economy, the heterogeneity of firm characteristics suggests a multiplicity of competitive strategies. Since the pioneering work of Penrose (1959) and Cyert and March (1963) this heterogeneity is thought to express the firm-specific assets that undergird resource-based visions of firm performance developed by Wernerfelt (1984) and Barney (1991). Kogut and Zander (1992) were among the first to emphasize the critical role of knowledge within this framework. What is clear from related empirical work is that firms search for efficiency and for knowledge in many different ways (Baily et al., 1992; Baldwin & Rafiquzzaman, 1995; Saxenian, 1994). Within economic geography, a standard distinction is made between those competitive advantages that are generated internally within the firm and those that emerge through co-location with other firms.

For the firms that agglomerate in space the collective resources that sustain the industrial district have long been envisioned, after Marshall (1920), as lower-cost access to specialized suppliers and buyers, to the associated pools of labor that clusters exploit and nurture, through to spillovers of knowledge. A somewhat different vision is offered by Jacobs (1969) who does not contest the efficiency of Marshall's (1920) districts, but who rather imagines the long-run prospects of firms to rest more squarely on the diversity that cities provide. A modern update is advanced by Duranton and Puga (2001). Glaeser et al. (1992) present empirical evidence of more rapid industrial growth within diversified local economies, while Henderson (2003) and Baldwin et al. (2010) provide firm-level evidence of higher levels of productivity in specialized urban economies. More recent work suggests that even within industrial districts the characteristics and behaviors of firms remain highly variable and that not all firms generate efficiencies in the same way (Neffke et al., 2011; Potter & Watts, 2011; Rigby & Brown, 2015).

For economic sociologists, these differences are explained by the structure of social networks that link firms and other economic agents (Burt, 2000; Powell, 1990). Social networks are broadly seen as an organizational form that enhances the sharing of knowledge and other resources in technologically complex industries where novel ideas are widely distributed and the rapidity of innovation generates considerable uncertainty (Hagedoorn, 1993; Powell et al., 1996). Though networks are commonly viewed as raising efficiency, precisely how firms are embedded within networks is critical to their performance (Granovetter, 1973; Uzzi, 1997). There is increasing evidence that networks with weak ties promote exploration and technological discovery, while networks with strong ties promote exploitation (Burt, 1992; Rowley et al., 2000; Walker et al., 1997). At the same time, the negative implications of over-embeddedness in networks are well known (Grabher, 1993).

Linking the concept of the industrial cluster with the relational perspective of social network analysis has generated a good deal of research that seeks to unpack the nature of collaboration and the mechanisms that guide the production and the flow of knowledge (see Gordon and McCann, 2000; Brenner et al., 2013). Boschma (2005) argues that understanding the interaction of different forms of proximity is critical to these efforts. Taking up this challenge, a series of empirical papers illustrate how the structure of social networks vary over space (Cantner and Graf, 2004; Cantner et al., 2010; Sorenson, 2005), how social proximity is shaped by other forms of distance (Bathelt et al., 2004; Knoben & Oerlemans, 2012; Malmberg & Maskell, 2006) and how networks evolve over space and time (Balland, 2012; Capello and Lenzi, 2013; Ter Wal, 2013). Cantner and Graf (2011) provide a useful overview.

While much of the work just mentioned takes the form of case studies, there is also a growing literature that explores the contribution of social networks to sustaining the agglomeration of knowledge production across systems of cities and regions. Singh (2005) shows that co-invention networks play a central role in the diffusion of ideas and knowledge amongst inventors. Several descriptive measures have been developed to characterize the structural aspects of regional co-inventor networks and used to explain the observed spatial variance in knowledge production. Bettencourt et al. (2007) find a significant positive relationship between the number of ties, the clustering coefficient, the size of the largest component among U.S. metropolitan co-inventor networks and the rate of patenting. Fleming et al. (2007) find that the size of the largest component (LC) of metropolitan inventor networks and the inverse path length between inventors is positively related to subsequent patenting in urban areas. Lobo & Strumsky (2008) report that inventor density (inventors per square mile), network aggregation and the ratio of non-local inventors in the metropolitan co-inventor network have a positive and significant relationship with the rate of metropolitan patenting. Unlike Fleming et al. (2007), they find a significant negative relationship between size of the LC and patenting. Strumsky and Thill (2013) examine the relationship between a series of metropolitan co-inventor network statistics and four metropolitan economic performance indicators (wage, income, jobs and GDP). Their results show that the relationships between these network statistics and metropolitan performance indicators are inconsistent, indicating that the nature of the relationship between network connectivity, knowledge production and regional economic performance is sensitive to the precise measures employed. Breschi & Lenzi (2016) explicitly attempt to measure the structure of internal and external co-inventor networks using the average inverse geodesic distance between any pair of linked inventors within an urban area. They find no significant relationship between greater internal or external social proximity and the rate of patent production. However, they find a positive and significant effect of the interaction between internal social proximity and clique density on the rate of patenting. Moreover, they report a positive and significant relationship between the interaction of internal and external social proximity and patent production.

While this research illustrates clearly how social network characteristics influence the pace of knowledge production within cities, it does not consider whether the architecture of regional knowledge stocks might shape the structure of social networks. The stocks of knowledge that accumulate in particular places may be characterized by their age and size, by their diversity across scientific, technological or industrial fields (Asheim et al., 2007; Kogler et al., 2013) and by their complexity (Balland & Rigby, 2017; Fleming & Sorenson, 2001). Is this variation correlated with the structure of innovation networks? Amin and Cohendet (2004), Moodysson (2008) and Trippl et al. (2009) make clear that the processes of innovation, the actors involved and the relationships between them vary significantly across different knowledge bases. Cantner et al. (2010) also suggest that more specialized regional knowledge bases are associated with larger and denser co-inventor networks. The empirical work that follows extends discussion of the linkages between the character of regional knowledge cores and the structure of local and non-local knowledge networks.

3. Data & Methods

The aim of this paper is to explain variations in knowledge production across U.S. metropolitan areas between 1975 and 2005, and to explore the roles of the structure of knowledge and social networks in such explanation. We measure knowledge production using patent data derived from the United States Patent and Trademark Office (USPTO). Our dependent variable is the annual number of patents produced within each U.S. metropolitan area. Note that we add a second dependent variable, patents per worker, as a robustness check in some of our output. Many patents are generated by more than one inventor. When these teams of inventors are located in the same metropolitan area, the individual patent is fully assigned to that location. In the case of patents produced by inventors located in different metropolitan areas, individual patents are fractionally split across those areas with shares determined by the geographical distribution of co-inventors. Patents developed solely by foreign inventors are excluded from our data. Fractional counts of patents imply that the dependent variable is not a “count variable”. Our fractional counts focus on the application year of patents as is customary in the literature. While patents provide a useful indicator of innovation, it should be clear that they are not a perfect measure (see Griliches, 1990).

Two independent variables play a central role in our analysis of metropolitan knowledge production. The first of these is a measure of urban co-inventor networks and the second is a measure of the specialization of a city’s knowledge core. We measure two social networks of co-inventors for each metropolitan area in each year, the first highlighting intra-city collaboration and the second built from inter-city linkages. Data gathered for additional covariates are identified next.

The USPTO lists the names of all inventors on patents. These inventors form the nodes of potential collaboration networks that vary year-by-year according to whether inventors have jointly applied for a patent in a given time-period. When two or more inventors are listed on the same patent then a link is established between the inventor-nodes. The addresses (city and county) are listed for all inventors on patents. We use the inventor county to assign individual patents (either fully or fractionally) to their corresponding CBSA[1]. The largest 366 CBSAs form the metropolitan statistical areas (MSAs) upon which our analysis is focused. When co-inventors on a patent are located in the same metropolitan area then we have an intra-city network link. When inventors on a patent are located in different metropolitan areas then we have an inter-city network link. Patents with more than two co-inventors can simultaneously represent intra- and inter-city network linkages. Both kinds of networks are examined below.

The number of nodes in our networks is given by the number of distinct inventors that we can link to patents. Unfortunately, the USPTO does not uniquely identify individual inventors. Thus, it is impossible to tell from USPTO records whether an inventor on patent *i*, in application year *t*, named John Smith is the same inventor as John Smith listed on patent *j* from the same application year *t*. To resolve such ambiguities, we utilize disambiguated inventor IDs made available by Lai et al. (2012) and link these to the inventors on all patents. Fleming et al. (2007) and Lobo and Strumsky (2008) proceed in the same way, identifying inventor networks over space and time using disambiguated inventor records.

However, comparing different metropolitan co-inventor networks is troublesome for two reasons. First, these networks tend to be disconnected. This means that within each urban area

[1] We use the December 2009 classification of CBSAs by the U.S. Bureau of the Census.

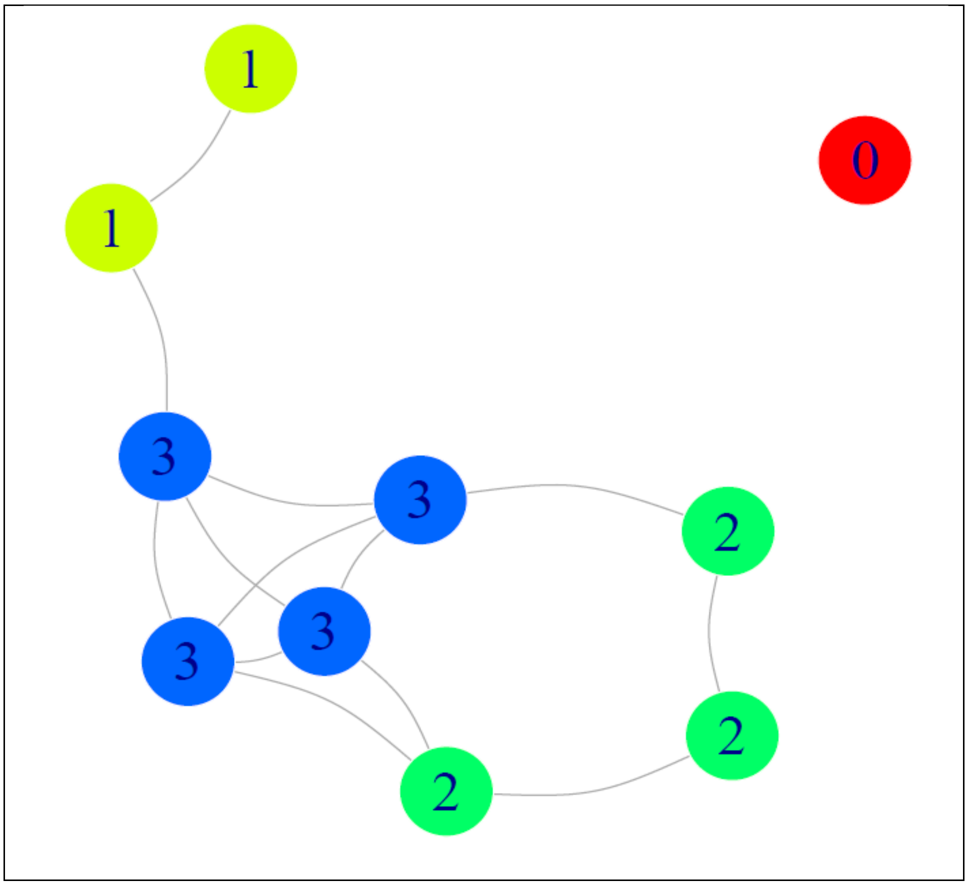
not every inventor is connected (directly or indirectly) to others through a co-inventor patent linkage. A lot of network-level measures, and especially centrality based measures, behave poorly and produce spurious results for disconnected networks (De Nooy et al., 2011). This feature makes most network measures unfit to use as covariates in a model to explain regional variations in knowledge production. Second, size and density are found to interact strongly with network measures. Networks with different sizes and/or densities can have significantly different probability distributions for the same network measure. Therefore, determining whether the observed value of the network measure is a direct result of structural network characteristics is difficult (Anderson et al., 1999). As a consequence, scholars interested in regional co-inventor networks have often limited their analyses to the largest component of networks, or used elementary descriptive statistics to characterize the co-inventor network. In both cases, the effects of co-inventor network structures on regional knowledge production are likely biased.

Fortunately, the k-core network measure developed by Seidman (1983) allows comparison of networks of different size and density and it is also applicable to disconnected networks (Butts et al., 2012). The k-core measure is a nodal degree based approach to identify cohesive subgroups. A k-core is a subgraph in which each node is connected to a minimum k other nodes in the subgraph (Seidman, 1983). Thus, k-core subgraphs contain nodes that have a specified number of ties to other nodes in the subgraph. Formally, a subgraph is a k-core when $d_s(i) \geq k$ for all $n_i \in N_s$, where $d_s(i)$ denotes the number of connections (*degree*) of every node n_i in the subset of vertices N_s, and k represents the order of the core. Matula and Beck (1983) developed an algorithm to decompose a full network into different k-cores. This algorithm identifies which nodes belong to 0-core but not to the 1-core or higher. These nodes are assigned to core number 0 and removed from the network for further consideration. Next, the algorithm identifies which nodes belong to 1-core but not to the 2-core or higher, etc. The algorithm stops when it is not able to identify nodes with links greater than the maximum k-core value. Networks with a larger number of k-cores tend to have greater variability in the connectedness of nodes than networks with a smaller number of k-cores, ceteris paribus. We use the number of k-cores to characterize the structure of inter- and intra-city co-inventor networks. We hypothesize that increasing the number of k-cores in inter- and intra-city co-inventor networks will result in greater metropolitan patent production.

In essence, the k-core measure captures the robustness of communities in a network under degeneracy. In networks that are more robust the connectivity amongst nodes is denser. In the context of our paper, this means that cities with a greater number of k-cores contain inventors that are more densely connected with one another. For example, consider the network in Figure 1. This network has ten inventors, nine that collaborate. The number printed in the nodes corresponds to the k-core the node belongs to. In this graph there are four different k-cores. The most robust community is the k-core with the greatest k value, the 3-core in this case. The inventors in this community each collaborate with at least three inventors that also collaborate with a minimum of three inventors. Caution is needed here, because the bipartite nature of our data (inventors on patents) can easily inflate this number. A patent with 16 co-inventors generates a k-core of 15. Similarly, an annual network in which two inventors collaborate on 22 patents produces a k-core of order 22, because each inventor has 22 linkages to a different inventor also with 22 linkages. Thus, using the size of the largest k-core is not so much a measure of the structure of the co-inventor network, as it is a measure of (repeated) collaboration amongst a selective subset of co-inventors (see Strumsky and Thill, 2013). Our preferred network measure counts the number of k-cores that occur in the city-level co-inventor network.

This measure captures the structure of the entire co-inventor network found either within a city (internal network) or between cities (external network).

Figure 1: An example of K-cores



There is a long history within economic geography that suggests the character of a city's knowledge base, one measure of industrial structure, might influence the pace of invention (Marshall, 1920; Jacobs, 1969). In particular, we are interested in whether metropolitan areas with more specialized or more diverse knowledge stocks generate more patents. Across much of the literature, the standard measure of specialization (or diversity) is the Herfindahl-Hirschman index (Hirschman, 1964). While this index is widely used, it has one major failing, namely its inability to control for varying "distances" between the economic categories across which specialization is measured. Here, we calculate the specialization (or diversity) of the knowledge base of cities by examining the distribution of patents across the 438 primary technological classes of the USPTO. For each pair of these classes we measure the technological distance or the cognitive proximity between them using patent co-classification data. We then compute the average relatedness or the average cognitive proximity between all pairs of patents that are generated within a city. This measure of average relatedness is bounded by the interval 0 – 1. Higher values of average relatedness indicate greater specialization.

The details of these calculations are outlined below. Co-class information on individual patents is employed to measure the technological proximity of technology classes, following the earlier work of Jaffe (1986), Engelsman & van Raan (1994), Kogler et al. (2013) and Nesta & Saviotti (2005). To measure the proximity, or knowledge relatedness, between patent technology classes in a single year we employ the following method. Let P indicate the total number of patent applications in the chosen year. Then, let $F_{ip} = 1$ if patent record p lists the classification code i, otherwise $F_{ip} = 0$. Note that i represents one of the 438 primary technology classes into which the new knowledge contained in patents is classified. In a given year, the total number of patents that list technology class i is given by $N_i = \sum_p F_{ip}$. In similar fashion, the number of individual patents that list the pair of co-classes *i* and *j* is identified by the count $N_{ij} = \sum_p F_{ip} F_{jp}$. Repeating this co-class count for all pairs of 438 patent classes yields the (438x438) symmetric technology class co-occurrence matrix ***C*** the elements of which are the co-class counts N_{ij}. The co-class counts measure the technological proximity of all patent class pairs, but they are also influenced by the number of patents found within each individual patent class N_i. Thus, we standardize the elements of the co-occurrence matrix by the square root of the product of the number of patents in the row and column classes of each element, or

$$S_{ij} = N_{ij} \Big/ \sqrt{N_i * N_j} \tag{1}$$

where S_{ij} is an element of the standardized co-occurrence matrix (***S***) that indicates the technological proximity, or knowledge relatedness, between all pairs of patent classes in a given year. The elements on the principal diagonal of ***S*** are set to 1. We prefer this simple form of standardization for the reasons outlined by Joo and Kim (2010).

The average relatedness value of patents for a metropolitan area *m* in year *t* is calculated as:

$$AR^{m,t} = \frac{\sum_i \sum_j S_{ij}^t * D_{ij}^{m,t} + \sum_i S_{ii}^t * 2D_{ii}^{m,t}}{N^{m,t} * (N^{m,t} - 1)} \qquad for\ i \neq j \tag{2}$$

where S_{ij}^t represents the technological relatedness between patents in technology classes *i* and *j*, $N^{m,t}$ is a count of the total number of patents in region *m* in year *t*, and where $D_{ij}^{m,t}$ counts the number of pairs of patents that can be located in technology classes *i* and *j* in region *m* in year *t*. To clarify the meaning of $D_{ij}^{m,t}$, imagine a region with three patents, one in technology class 1 and two in technology class 2. Then, the pair counts $D_{ij}^{m,t}$ represent elements in the (438x438) symmetric matrix

$$\boldsymbol{D}^{t,r} = \begin{bmatrix} 0 & 2 & \dots & 0 \\ 2 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \quad (3)$$

with three patents, there are 3x2 = 6 unique distance measures to calculate, the distance between the patent in class 1 and each of the patents in class 2, the distances from both patents in class 2 to the patent in class 1 and the distance between the two patents in class 2. Note that the latter distance is counted twice. These routines are repeated for each of the 35 years in our analysis. Descriptive statistics for the average relatedness variable are provided in Table 1.

Cities and regions that build knowledge stocks around particular industries and technologies will likely record different numbers of patents over time as some sectors of the economy heat up and others cool down. Patents generated in very dynamic technology classes likely build incrementally on recent patents in the same sector. One way of controlling for the distribution of urban knowledge stocks across more or less dynamic classes is to capture the average age of citations on the patents generated each year. Cities active in newer technologies will likely have citations that are more recent than cities where invention is in older technologies. As patents are indexed by USPTO numbers that track the timing of their introduction to the economy, we calculate the mean age of citations on patents by averaging the USPTO numbers of the patents that they cite. When this average number is higher it references recent patents or newer technologies. We anticipate that metropolitan areas that are over represented in newer technologies will thus cite patents that have higher USPTO numbers on average. Including this mean age of citations should control for the degree to which urban areas are active in more dynamic technological sectors. Other authors in this field have used similar approaches (Fleming et al. 2007; Strumsky and Thill 2013; Breschi and Lenzi 2016).

Cities that devote significant resources to invention are more likely to produce more patents than cities that don't make such investments. Typically research and development (R&D) spending and venture capital funding are obvious indicators of such efforts. Unfortunately, there are no R&D data available at the city-level that cover the period of our investigation. However, we can make use of venture capital data and National Science Foundation (NSF) funding data to proxy for the local availability of funds for knowledge production. The venture capital data originate from the Thomson VentureXpert series (see Samila and Sorenson, 2011), and the NSF funding data can be found on the NSF website. These data can be readily aggregated to the MSA level for individual years[2]. We combine the venture capital data and the NSF awards by metropolitan area as an indicator of R&D spending across U.S. urban areas. We hypothesize that higher levels of R&D funding should be associated with a larger volume of patents.

The level of competition within the metropolitan region might affect inventive activity. There is significant disagreement as to whether local monopoly (MAR-model) or competition (Porter, 1990) foster inventive activity. On the one hand, the monopoly argument holds that firms with market power are more likely to invent because they can fully appropriate the economic benefits from their efforts. On the other hand, the argument for competition holds that a firms' inventive activity benefits from knowledge externalities (Audretsch & Feldman, 1996). We control for the level of economic competition within a metropolitan area by calculating the ratio of the number of firms to employment. Higher levels of this ratio signify greater competition. Counts of the number of firms and employment at the county level may be found in the County Business Patterns data generated by the U.S. Bureau of the Census. County figures are summed across the regional units that comprise each MSA. We have no explicit hypothesis

[2] Data available at https://www.nsf.gov/awardsearch/download.jsp

on how competition impacts knowledge production, reflecting ambiguity in the existing literature.

Clearly larger MSAs are expected to generate more patents than smaller MSAs. Employment within urban areas, obtained from County Business Patterns (U.S. Census Bureau), is used to control for urban scale or size effects and for the higher levels of interaction that scale facilitates. The density of inventors (inventors/land area) is a proxy for the level of MSA agglomeration. We hypothesize that larger cities and cities with higher levels of inventor density will exhibit higher levels of patenting. Figure 2 shows the spatial distribution of the number of patents for the period 2006-2010 across U.S. metropolitan areas.

Descriptive statistics for all variables are shown in Table 1 for four time periods spanning most of the period under investigation. Approximately 2.3 million patents were generated in the 366 U.S. metro areas over the period 1976 to 2010. The New York MSA produced most patents since 1975 accounting for 166,000 of the total. In second place, San Jose inventors produced approximately 153,000 patents over the study period. In third place, Los Angeles inventors generated approximately 117,000 patents. San Francisco, Chicago and Boston occupy the next three ranks in terms of urban knowledge production since 1975. Hinesville-Fort Stewart, GA produced the fewest patents (15) of any metropolitan area over the period examined.

Note that the relatively large values for the average age of citations in Table 1 reflects the fact that we calculate the mean citation age of patents within an urban area by examining the USPTO numbers on all patents that are cited by inventors in a particular city and year. Utility patent issue numbers start at 3858241 in 1975. Thus, for 1980, the average age of citations (3732129) corresponds to an average date of issue of 1973 (an average age of 7 years). The average relatedness value (index of knowledge specialization) across U.S. metropolitan areas was 0.026 in 1980, increasing to 0.043 in 2010. Knowledge production is becoming more specialized at the urban level across the United States. This means that the average “technological distance” between all pairs of patents generated within a metropolitan area is declining over time. Relatively high values of the average relatedness variable, approaching the maximum value of 1, are usually associated with small cities that produce only a few patents concentrated in one or two classes.

Total R&D data are listed in millions of dollars in Table 1. In any specific year these figures exhibit considerable left hand (negative) skew as only a small number of cities capture the largest volumes of venture capital. NSF funding tends to be much less volatile than venture capital investments and more broadly distributed. The mean values reported for R&D are inflated by the extreme values in the venture capital series. Note also that the venture capital investments are extremely volatile year-to-year. The maximum value reported in 1990 was approximately one-tenth the maximum in 1991, reflecting the sharp business cycle downturn of that year. Yet, only 10 years later in 2000, venture capital investment across the U.S. expanded to more than $100 billion, just as the dot.com bubble collapsed.

Table 2 illustrates correlation coefficients between the variables, prior to taking logarithms. While the Pearson correlation coefficients are reasonably large in a few cases, the coefficients in models with/without core variables are relatively stable. The reader is reminded that multicollinearity does not bias estimators, it merely makes then inefficient (Goldberger, 1991). Inefficiency does not appear to be a problem in the results presented below.

Table 1: Descriptive Statistics

Variables	**1980**				**1990**			
	Mean	**S.D.**	**Min**	**Max**	**Mean**	**S.D.**	**Min**	**Max**
Patents	97.84	303.59	0	3780.14	135.71	373.77	0	4030.8
Employment	258919	643811	4214	8250921	319663.9	770983	8758	9344755
Inventor density	0.02	0.05	0	0.59	0.03	0.06	0	0.64
Ave. relatedness	0.02	0.03	0	0.21	0.03	0.05	0	0.82
Citation age patent	3732129	449067	0	4215443	4253089	340550	0	4685821
Total R&D	27.41	265.24	0	3700.56	15.83	75	0	828.06
Firms per Emp.	0.04	0.01	0.01	0.06	0.04	0.01	0.02	0.08
Internal k-cores	2.98	2.43	0	26	4.09	2.89	0	19
External k-cores	1.96	1.63	0	10	3.07	2.44	0	13

Variables	**2000**				**2010**			
	Mean	**S.D.**	**Min**	**Max**	**Mean**	**S.D.**	**Min**	**Max**
Patents	270.14	803.13	1.5	8478.3	236.79	749.48	0	8373.63
Employment	383428	870740	14277	10099328	407475	925715	22830	11255588
Inventor density	0.06	0.14	0	1.48	0.06	0.15	0	1.72
Ave. relatedness	0.03	0.03	0	0.51	0.04	0.07	0	0.71
Citation age patent	5257239	168177	4720271	5720329	5607179	455949	4077948	6782696
Total R&D	232.22	1562.67	0	21006.58	32.18	167.47	0	1781.93
Firms per Emp.	0.04	0.02	0	0.32	0.04	0.01	0.01	0.08
Internal k-cores	6.27	5.37	0	32	6.1	5.63	0	34
External k-cores	5.95	5.42	0	37	6.01	5.91	0	44

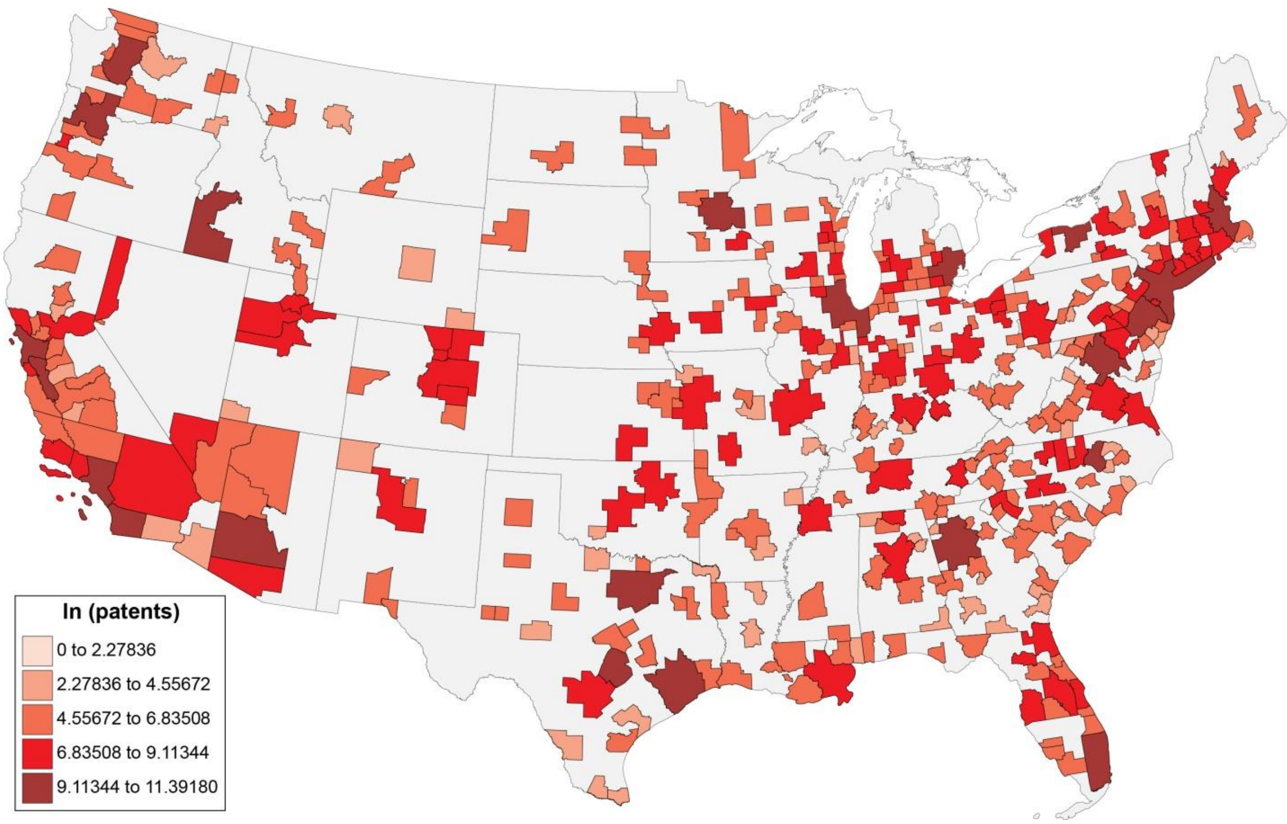
Notes: The variables in Table 1 are not logged.

Table 2: Correlation Among Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Patents (1)	1								
Employment (2)	0.78	1							
Inventor density (3)	0.77	0.44	1						
Average relatedness (4)	0	-0.05	0.01	1					
Citation age patents (5)	0.17	0.12	0.22	0.11	1				
Research & Dvlpmnt (6)	0.6	0.45	0.53	-0.01	0.07	1			
Firms per employment (7)	-0.03	-0.03	-0.03	0.01	0.01	-0.02	1		
Internal K-cores (8)	0.75	0.63	0.65	0.1	0.39	0.37	-0.05	1	
External K-cores (9)	0.72	0.57	0.71	0.08	0.45	0.4	-0.04	0.9	1

Notes: The correlations are based on the raw values of variables, not logged values.

Figure 2: The spatial distribution of the number of patents for the period 2006-2010 across U.S. metropolitan areas



4. Results

We anticipate that knowledge production in U.S. cities might be influenced by inventive activity taking place in other parts of the U.S. urban system. Indeed, statistical tests revealed significant positive spatial autocorrelation in MSA patent output. The geography of U.S. cities was represented in these tests with an inverse distance matrix generated for all city-city pairs. In the regression models presented in Tables 3 and 4, a spatial lag term controls for spatial autocorrelation using the *spdep* and *splm* packages in R. Estimation makes use of fixed effect panel models covering 35 years and 366 metro regions. These models control for unobserved variables that are fixed at the MSA level. We control for time-specific shocks by adding time fixed-effects to our model, and as a crude control for concerns with endogeneity all independent variables are lagged by one-period. We control for heteroscedasticity using White's robust standard errors.

Table 3 presents the first results exploring whether cities that are specialized or diversified in terms of knowledge production produce more patents. Model 1 in Table 3 is offered as a baseline, ignoring concerns with spatial autocorrelation and not including co-inventor networks. The independent variables included in model 1 function largely as hypothesized. We control for the influence of MSA size with the employment variable. Not surprisingly, larger urban areas with higher levels of employment on average generate significantly more patents than smaller urban centers. Our simple measure of the strength of agglomeration within urban areas is inventor density. Increases in density raise the number of patents produced, as hypothesized. The age of technology has no significant influence on knowledge production in model 1. In line with most studies of knowledge production, as R&D investments in a city increase, inventive output also increases. Most importantly, perhaps, the average relatedness variable is significant and has a positive sign suggesting that more specialized cities tend to produce more patents. Our measure of competition, the number of firms in an MSA per worker, is insignificant in model 1[3]. It is important to note that the more traditional measures of urban agglomeration, employment and inventor density, have coefficients that are considerably larger than the social network measures. Given the log-log form of the model these may be interpreted as elasticities.

Adding spatial autocorrelation in model 2 revealed that both spatial lag and error terms in the autocorrelation model were significant. Lagrangian multiplier tests suggested the lag form of autocorrelation was stronger and so spatial lags were added to all models. A comparison of models 1 and 2 indicates that many of the independent variables have similar coefficients after the introduction of the spatial lag term. There are two exceptions. First, the citation age of patents, a measure of the age of technology becomes significant after the introduction of the spatial lag term. Second, the influence of R&D spending becomes insignificant in models including controls for spatial autocorrelation. This is not that surprising, as Samila and Sorenson (2011) report quite mixed results for the influence of venture capital funding on measures of metropolitan economic performance. We do not report the coefficient of determination as it is inflated by the spatial lag term.

Models 3-5 introduce network measures to our analysis of urban knowledge production. Like Strumsky and Thill (2013), we capture the structure of internal and external city networks using k-core counts, though we exploit the k-core measure a little differently as indicated above. In line with existing studies (Fleming et al. (2007), Lobo and Strumsky (2008) and Breschi and Lenzi (2016)), model 3 shows that the structure of co-inventor networks, those that are internal

[3] Model 1 has fewer observations than our other models, because it is the only model that is fitted with OLS. This technique doesn't deal well with unbalanced panel data, hence incomplete cases are dropped.

to the city and those that link collaborators within a city to inventors elsewhere ("external networks"), have a positive and significant influence on patent production within the city. Indeed, denser webs of collaboration amongst inventors (either internal or external) foster the production of patents. These network effects are initially treated as independent of our measure of urban agglomeration that is captured by inventor density. Note that the internal network measure has a stronger influence on knowledge production than the external network measure. It seems reasonable to anticipate some interaction between the measures of agglomeration and co-inventor networks (see Whittington et al., 2009). This concern is the focus of models 4-6. Thus in model 4, we interact inventor density (our measure of urban agglomeration) with the number of internal co-inventor k-cores in the city to examine whether or not internal collaboration networks are a complement or a substitute for agglomeration. The negative coefficient on the internal interaction variable in model 4 indicates a substitution effect and suggests that cities with large urban agglomerations gain less from local networks than cities where such agglomeration is rather poorly developed. Model 5 supports a similar story of substitution between the forces of agglomeration within cities and external collaboration networks. These results are somewhat surprising. Our intuition led us to suspect that at the city-level, agglomeration and networks would act as complements, combining to raise the overall volume of knowledge production, especially in the case of external knowledge networks. Perhaps the spatial lag term dampens the influence of the external network measure.

Model 6 explores the interaction between internal and external knowledge networks at the city-level. The negative coefficient for the interaction variable indicates that the number of internal and external k-cores act as substitutes. This doesn't make sense intuitively on the city-level, where we might expect that external collaborations (pipelines) feed the internal inventor

Table 3: Determinants of the Pace of Patenting in U.S. Metropolitan Areas

Dependent variable: ln patents	(1)	(2)	(3)	(4)	(5)	(6)
Spatial Autocorr.		0.545***	0.520***	0.596***	0.568***	0.527***
		(0.042)	(0.043)	(0.039)	(0.041)	(0.043)
Employment	1.223***	1.075***	0.848***	0.804***	0.915***	0.837***
	(0.028)	(0.149)	(0.136)	(0.134)	(0.141)	(0.136)
Inventor density	3.037***	3.325***	2.169***	6.101***	6.088***	3.112***
	(0.100)	(0.541)	(0.503)	(1.077)	(1.051)	(0.695)
Ave. relatedness	1.688***	2.701***	2.017***	1.995***	2.316***	2.020***
	(0.099)	(0.532)	(0.484)	(0.476)	(0.502)	(0.481)
Citation age patent	0.001	0.053***	0.042***	0.042***	0.051***	0.041***
	(0.003)	(0.015)	(0.013)	(0.013)	(0.014)	(0.013)
Research & Development	0.015***	0.009	0.002	-0.001	-0.003	0.004
	(0.005)	(0.026)	(0.024)	(0.023)	(0.025)	(0.024)
Firms per Emp.	0.061	-0.118	0.005	-0.067	-0.133	-0.009
	(0.219)	(1.210)	(1.087)	(1.070)	(1.134)	(1.082)
Internal K-cores			0.486***	0.568***		0.502***
			(0.071)	(0.060)		(0.071)
External K-cores			0.127**		0.344***	0.148**
			(0.062)		(0.056)	(0.062)
Interaction internal				-0.135***		
				(0.034)		
Interaction external					-0.104***	
					(0.028)	
Interaction int. * ext.						-0.001*
						(0.0004)
N	12,444	12,810	12,810	12,810	12,810	12,810
CBSA	366	366	366	366	366	366

Notes: *p < .1 **p < .05 ***p < .01

The independent variables are log transformed with the exception of the citation age of patents. Excluding the spatial lag, the independent variables are lagged one period. Year fixed effects included but not shown. The model is a fixed effects panel with robust standard errors. The terms in parentheses at the top of the table represent different models discussed in the text.

pool (local buzz) with non-local knowledge thus boosting overall knowledge production (after Bathelt et al., 2004). Perhaps it is the case that inventors can only collaborate with a finite number of partners. Hence, collaborating with co-inventors located in other metropolitan areas limits the opportunities to collaborate within the city and vice versa (see Esposito and Rigby, 2018). Note that these findings are inconsistent with the results of Breschi & Lenzi (2016) who report a positive interaction between internal and external network effects on urban invention[4].

We now shift to examine the influence of co-inventor networks on knowledge production in metropolitan areas that are relatively specialized or relatively diverse in terms of the range of technologies that they contain. Figures 3 and 4 illustrate how co-inventor networks vary across metropolitan knowledge cores. Table 4 summarizes this variation, reporting descriptive statistics on k-core values for specialized and diversified urban areas. Overall, specialized cities tend to have much more well-developed internal and external co-inventor networks than diversified metropolitan areas and this finding holds for cities of different size. For example, Figure 3 clearly shows differences in the internal collaboration network structure of the medium-sized cities Boise and Pittsburgh in 2005. On average, inventors in Boise, a specialized city, are much more connected to other local inventors than inventors in a more diversified city such as Pittsburgh. In larger cities we see the same pattern, with a much more well-developed internal network in San Jose, a specialized metropolitan area, than in Chicago which is technologically more diversified. Figure 4 illustrates these same differences for external co-invention networks in the smaller and larger MSAs of Poughkeepsie and Cleveland and again in San Jose and Chicago.

[4] Note that limiting our analysis to census years, for which we have educational data, and including the share of MSA population with a bachelor's degree as a measure of human capital, analysis produces broadly similar results to those reported in Table 1 and yields a positive coefficient on the human capital variable.

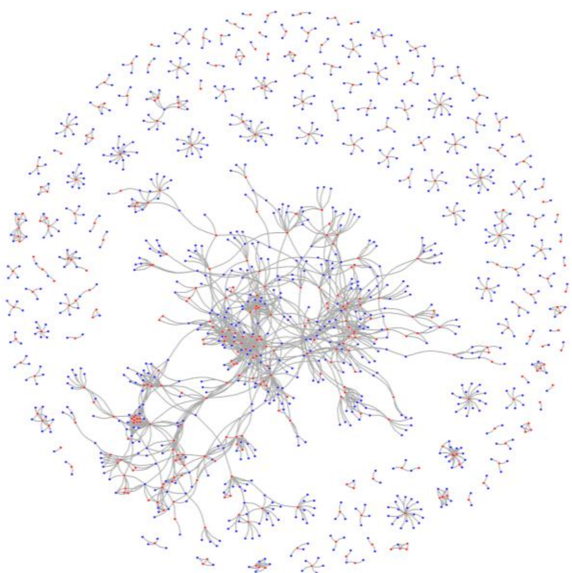
Figure 3: Internal (within-city) Co-Inventor Networks in Specialized and Diversified Urban Areas

Specialized Metropolitan Areas	Diversified Metropolitan Areas

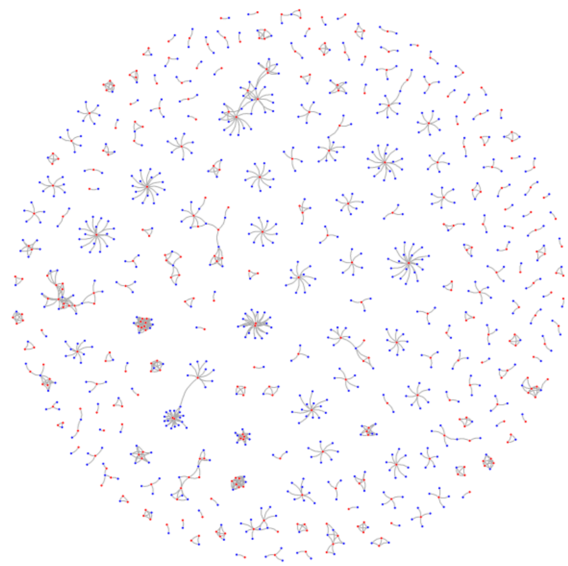
Figure 4: External (between-city) Co-Inventor Networks in Specialized and Diversified Urban Areas

Specialized Metropolitan Areas	Diversified Metropolitan Areas

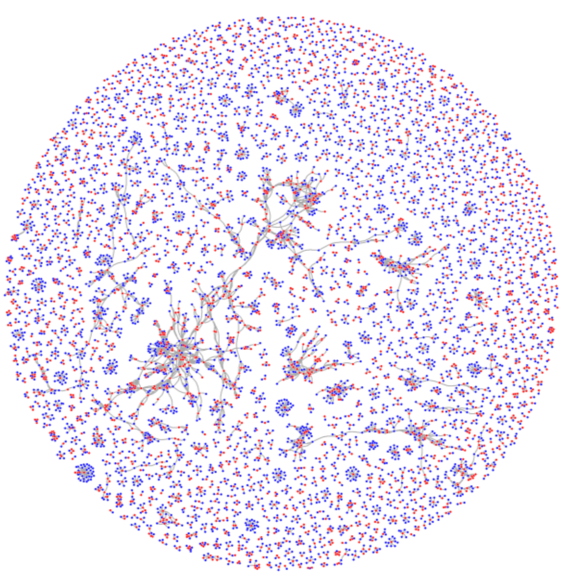
External inventor network Poughkeepsie (NY) 2005



External inventor network Cleveland (OH) 2005



External inventor network San Jose (CA) 2005



External inventor network Chicago (IL) 2005

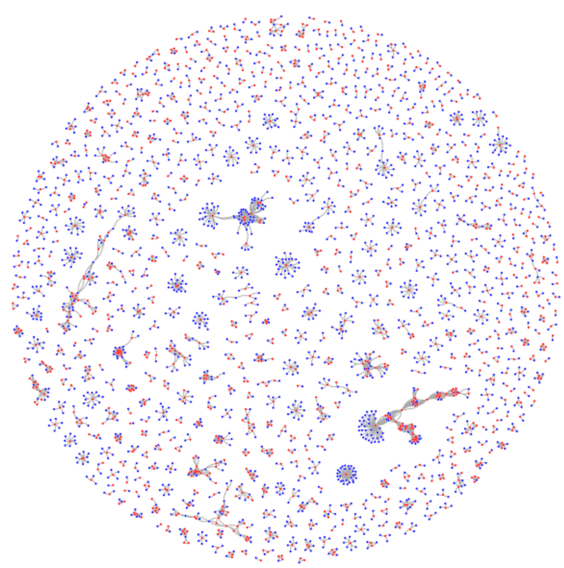


Table 4 summarizes data on the characteristics of urban co-inventor networks in specialized and diversified cities across the four years that span the period of analysis. The table highlights descriptive statistics on the number of internal and external k-cores and reports the maximum k-core values. The statistics are shown for cities in the top and bottom quartiles of the distribution of average relatedness (technological specialization) values. Cities in the lower quartile have more diversified knowledge cores. Cities in the upper quartile have specialized knowledge cores. The lower quartile contains a few less observations than the upper quartile because of metropolitan areas with zero patents (no average relatedness scores). The allocation of cities to the quartiles is done year-by-year. Observations increase over time as invention spreads across more U.S. cities. In general, the mean (and maximum) number of k-cores, internal and external, is greater in cities with specialized technological cores than in cities with more diversified knowledge bases.

Figures 3 and 4, along with Table 4, suggest differences in the structure of co-inventor networks between specialized and diversified urban areas. To examine whether the structure of co-inventor networks operate differently in cities with specialized knowledge cores and those that are more diversified, we run the same regressions as in Table 3 across the two groups of cities. The results are presented in Table 5.

Table 4: Co-inventor Network Statistics for Cities with Specialized and Diversified Knowledge Stocks

	Internal Network k-cores				External Network k-cores			
	Mean	S.D.	Max	# Obs.	Mean	S.D.	Max	# Obs.
1980								
Specialized	4.262	1.547	9	42	2.690	1.199	6	42
Diversified	4.115	2.123	11	26	2.692	1.761	6	26
1990								
Specialized	5.697	2.744	15	33	4.152	1.873	9	33
Diversified	4.929	2.761	13	28	4.071	2.581	12	28
2000								
Specialized	9.588	5.231	25	51	8.890	5.591	25	51
Diversified	7.417	5.369	28	36	7.083	4.735	22	36
2010								
Specialized	6.554	5.696	33	92	6.011	5.841	43	92
Diversified	2.065	1.212	8	92	2.152	1.617	12	92

The regression models that underpin the results in Table 5 are panel models estimated with ordinary least squares techniques. In all models, year fixed effects are included but not reported in the results. We are unable to control explicitly for spatial autocorrelation in the models of Table 5 because of the unbalanced nature of our panel data following its separation into specialized and diversified city-time components. As a proxy for spatial autocorrelation we add another variable to the models that represents a spatial lag term, measuring the inverse distance weighted value of patents generated in all cities save for the focal MSA. This spatial lag term is positive and significant and appears to operate much like the lag term in the models with spatial autocorrelation in Table 3. Note that we estimate four models in Table 5. Models 7 and 8 have the log of patents in the city as the dependent variable, following Table 3, while models 9 and 10 have the log of patents per worker as the dependent variable. We show models 9 and 10 just to indicate that the results are qualitatively the same with a slightly different specification of the dependent variable. We do not discuss models 9 and 10 further.

Model 7 in Table 5 reports the coefficients for our standard model of knowledge production for the set of cities that are diversified. The volume of employment has a positive and significant impact on knowledge production in cities with diversified knowledge cores. However, the remaining independent variables in model 7 are insignificant, including traditional measures of agglomeration (inventor density) and the k-core measures of internal and external social networks. As we switch to technologically specialized cities in model 8, all independent variables have a positive and significant impact on knowledge production except R&D and the number of firms per employee. Most importantly, the number of k-cores for internal and external co-inventor networks exerts a significant positive influence on knowledge production for specialized cities, unlike the finding for diversified cities.

Table 5: Knowledge Production in Specialized and Diversified Cities

Dependent variables: ln patents / ln patents per worker	(7) *Diversified Cities*	(8) *Specialized Cities*	(9) *Diversified Cities*	(10) *Specialized Cities*
Spatial Lag	0.408***	0.682***	0.034***	0.058**
	(0.116)	(0.282)	(0.009)	(0.025)
Employment	1.039***	0.887***	0.057***	0.043***
	(0.205)	(0.153)	(0.018)	(0.013)
Inventor density	0.957	3.712***	0.017	0.289***
	(0.669)	(1.120)	(0.061)	(0.099)
Ave. relatedness	-0.521	1.314***	-0.048	0.109***
	(0.917)	(0.287)	(0.082)	(0.023)
Citation age patents	7.87E-08	3.98E-07***	7.30E-09	3.33E-08***
	(4.99E-08)	(8.96E-08)	(4.56E-09)	(7.73E-09)
R&D	0.004	-0.001	0.000	-0.000
	(0.004)	(0.002)	(0.000)	(0.000)
Firms per Emp.	5.487	2.022	0.408	0.151
	(4.682)	(1.764)	(0.414)	(0.159)
Internal K-cores	0.043	0.257***	0.003	0.0212***
	(0.054)	(0.054)	(0.005)	(0.004)
External K-cores	-0.027	0.074*	-0.003	0.006*
	(0.049)	(0.041)	(0.004)	(0.003)
N	1,081	1,619	1,081	1,619

Notes *p < .1 **p < .05 ***p < .01

The independent variables are log transformed with the exception of the citation age of patents. Excluding the spatial lag, the independent variables are lagged one period. Year fixed effects are included but not shown. Standard errors are robust. The terms in parentheses at the top of the table represent models discussed in the text.

The results in Table 5 confirm that the nature and importance of co-inventor collaboration networks vary with the technological profiles of urban areas. We suspect that in diversified knowledge cities the breadth of the cognitive overlap between groups of inventors is not sufficiently high for dense networks of collaborating agents to form. In contrast, specialized cities channel knowledge development along relatively narrow trajectories that engender greater cognitive overlap and more readily hasten a shared division of labor in the knowledge production process. In turn, the efficiency of greater specialization and interaction sustain higher levels of knowledge output in cities with higher levels of cognitive proximity among inventors. Though our data support this notion, clearly more work is required to bolster this claim.

Finally, we examine whether inventors in specialized cities develop external (between city) collaborations in ways that differ from inventors located in diversified cities. Again we use the relatedness of a city's knowledge stock to classify metropolitan areas as specialized (average relatedness between a city's patents is in the upper quartile) or diversified (average relatedness between a city's patents is in the lower quartile). We then explore the factors that encourage collaboration between pairs of cities, or more accurately between inventors located in different pairs of cities. In our regression framework, the dependent variable takes the value 1 (0) when inventors in one city do (do not) collaborate with inventors in a second city. Our model of collaboration rests on a simple gravity framework where we anticipate that the probability of an external collaboration between a pair of cities is a positive function of the number of inventors in each city and a negative function of the geographical distance between them. We add to this simple specification a variable that captures the cognitive proximity between each pair of cities, measured as the average relatedness between all patents produced in the two cities in a given year. We hypothesize that as the cognitive proximity between cities increases, so inventors in those cities are more likely to collaborate.

To ensure that we don't count city-city pairings twice our observation set is cut in half. A dummy variable is added to the right-hand side of the regression model, taking the value 0 (1) when the observation refers to a collaboration originating in a diversified (specialized) city. We interact all independent variables with this city dummy to test whether the influence of those variables differs between diversified and specialized cities. With a binary dependent variable, we estimate these effects with a logit model fitted using maximum likelihood techniques.

Table 6 reports the results. The coefficients in the logit model are to be read as the log odds of the probability of collaboration between a pair of cities. When focusing on diversified cities (the specialization dummy takes the value zero), the probability of inter-city collaboration increases as the number of inventors in a pair of cities increases and as the distance separating them falls. These results are just as we might expect. In addition, as the technological profiles of the cities become more similar, as their cognitive proximity increases, then collaboration between a pair of inventors located in each of the two cities is more likely. As the index of city specialization turns to 1, we see that specialized cities in general engage in significantly less collaboration than their diversified partners, at least in terms of city-to-city linkages. The interactions in the model reveal that as the size of cities increases, the effect on the probability of external inventor collaboration is significantly lower in specialized cities than in diversified cities. This might be read as suggesting that size alone is a less important factor for collaboration in specialized as compared to diversified cities. The positive coefficient on the interaction of geographic distance and specialization indicates that inventors in specialized cities are less

Table 6: Metropolitan Collaboration in Diversified and Specialized Cities

Dependent variable: Collaboration (0/1)	
Dummy: Specialization	- .37088***
	(.01532)
# Inventor city *i*	.95554***
	(.00470)
# Inventor city *j*	.96550***
	(.00514)
Geographical distance	- 1.01131***
	(.00951)
Cognitive proximity	38.69149***
	(1.22060)
Interact. dummy * # inventors city *i*	- .05488***
	(.00611)
Interact. dummy * # inventors city *j*	- .04969***
	(.00584)
Interact. dummy * geographical distance	.17159***
	(.01547)
Interact. dummy * cognitive proximity	17.85297***
	(1.60817)
N	1202648
Prob. > Chi^2	0.0000
Pseudo R-Squared	0.4527

Notes: *p < .1 **p < .05 ***p < .01

All independent variables are logged.

Year fixed effects included but not shown.

impacted than inventors in diversified cities by increases in the distance separating them from potential collaborators. Finally, the positive coefficient on the interaction between cognitive proximity and specialized cities shows that technological relatedness is more important to inventors in specialized cities when forming their external collaborations than it is for inventors in diversified metropolitan areas. These findings are robust when running a linear probability model and when explicitly estimating the number of between city collaborations in a negative binomial specification.

Overall, these results establish that the forces influencing between-city tie-formation are significantly different for inventors in specialized urban areas and those in diversified urban areas. Tie formation across all cities is a positive function of the size of potential interacting partner cities, a positive function of the similarity of the knowledge base across cities and a negative function of the geographical distance between them. However, inventors in specialized cities are more selective than inventors in diversified cities when it comes to choosing their partners. They are more likely to engage with co-inventors in other cities when those partners exhibit greater technological similarity and they are less dissuaded by the friction of distance when doing so. The size of interacting partner cities is significantly less important for inventors in specialized cities than inventors in diversified cities.

5. Conclusion

Knowledge production is concentrated in cities where the density of economic agents is relatively high. That density encourages interaction and fuels processes of agglomeration that reinforce urban advantage, at least for some economic agents. Where clusters of firms and other economic actors combine to form social networks the pace of invention is also accelerated. In general, internal city networks exert a stronger influence on the pace of urban invention than external networks that link co-inventors across cities. Both internal and external co-inventor networks act as substitutes for agglomeration or the positive influence of inventor density on the pace of knowledge production. Internal and external networks also substitute for one another.

We show that the influence of social networks on urban invention is strongly conditioned by the architecture of knowledge found within cities. We capture that architecture with a simple measure of metropolitan knowledge cohesion that captures the average relatedness, or the technological proximity, between the patents produced in a city. Within specialized cities average relatedness or technological proximity is higher than in diversified cities. Metropolitan areas with specialized knowledge cores tend to be associated with more robust or denser social networks of co-inventors that add significantly and positively to the pace of invention. This is true for both internal social networks and external social networks. Metropolitan areas with diversified knowledge cores have social networks that are much less well-developed than specialized cities and which have no significant impact on the rate of invention.

Finally, we report that the social ties linking co-inventors across cities are shaped by the technological characteristics of the cities in which inventors reside. Inventors in metropolitan areas that have specialized knowledge cores are significantly less constrained by geographical proximity and significantly more constrained by cognitive proximity in their search for collaborators than are inventors located in urban areas with diversified knowledge cores. The pipelines that connect diversified cities appear shorter and less focused in terms of technology.

As we shift to consider the policy impacts of our results, like Lobo and Srumsky (2008), we find that traditional measures of agglomeration such as inventor density have a more significant influence on the pace of invention than the organization of social networks. In short, cities might do better to rely on the broadly defined economies associated with size and density rather than worrying about the structure of partnerships between inventors, either close or far. This is not so say that networks are unimportant, for we might presume that the benefits derived from agglomeration are, at least in part, built from interactions of various sorts. It does suggest, however, that much more work is required to specify the nature of networks and types of interactions found in cities, and to unpack the characteristics of these structures that are most important for the economic processes under investigation.

While these broad claims speak to cities as a whole, it should be clear from the findings above that cities with specialized and diversified knowledge cores pose rather different sets of policy concerns. Metropolitan areas that have diversified knowledge cores, with patents that are widely distributed across technology fields, exhibit variable levels of inventive productivity that are unrelated to the structure of their social networks, as measured by the number of k-cores, and unrelated to inventor density. Indeed, the primary driver of invention in diversified urban areas is city size. While the size of the city has long been employed as a proxy for urbanization economies, it is a proxy that offers few insights into the processes that may be at work. For Jacobs (1969), size was explicitly linked to diversity and to the possibilities of novel recombination. Given the nature of the relatedness measure that we employ above, those possibilities would seem to hinge around the concept of unrelated variety and the potential for breakthrough invention. We know little at this time about how to foster such forms of creativity.

In urban areas with specialized knowledge cores and well-established social networks, the relationship between agglomeration and social networks remains unclear. Like Whittington et al. (2009), we find that inventor density interacts negatively with our social network measure, indicating that agglomeration and social networks are substitutes rather than complements in the process of invention. Cities, then, have some flexibility in creating the conditions under which the production of ideas might be enhanced. Unlike Breschi and Lenzi (2016), we find that internal and external city co-inventor networks do not interact positively. Thus, while urban knowledge production benefits from more robust internal and external inventor networks, once those networks are operational, further development of one of these networks does not translate into additional gains from the other. This finding runs counter to the claims that non-local linkages necessarily enhance the benefits of local networks. One possible explanation for this is that local and non-local interactions might have different characteristics that are not complementary. For instance, Van der Wouden (2018) reports that patents generated by local collaboration tend to be more complex than patents produced by non-local collaboration. Much more work is clearly required to understand the heterogeneity in patterns of urban invention and interaction, and how the factors that shape the geography of knowledge creation evolve over time.

References

Agrawal, A., Cockburn, I., & McHale, J. (2006). Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography*, *6*(5), 571–591.

Amin, A., & Cohendet, P. (2004). *Architectures of Knowledge, Firms, Capabilities and Communities.* Oxford: Oxford University Press.

Anderson, B. S., Butts, C., & Carley, K. (1999). The interaction of size and density with graph-level indices. *Social Networks*, *21*(3), 239–267.

Arthur, W. B. (1999). Complexity and the economy. *Science*, *284*(5411), 107–109.

Asheim, B., Coenen, L. and J. Vang (2007). Face-to-face, buzz and knowledge bases: sociospatial implications for learning, innovation and innovation policy. *Environment and Planning C: Government and Policy* 25, 655-670.

Audretsch, D. B., & Feldman, M. P. (1996). R & D Spillovers and the Geography of Innovation and Production. *The American Economic Review*, *86*(3), 630–640.

Baily, M. N., Hulten, C., Campbell, D., Bresnahan, T., & Caves, R. E. (1992). Productivity Dynamics in Manufacturing Plants. *Brookings Papers on Economic Activity. Microeconomics*, *1992*, 187–267.

Baldwin, J. R., Brown, W. M., & Rigby, D. L. (2010). Agglomeration economies: Microdata panel estimates from Canadian manufacturing. *Journal of Regional Science*, *50*(5), 915–934.

Baldwin, J. R., & Rafiquzzaman, M. (1995). Selection versus evolutionary adaptation: Learning and post-entry performance. *International Journal of Industrial Organization*, *13*(4), 501–522.

Balland, P.-A. (2012) Proximity and the evolution of collaboration networks: Evidence from research and development projects within the global navigation satellite system (GNSS) industry. *Regional Studies* 46, 741-746.

Balland, P.-A., & Rigby, D. (2017). The Geography of Complex Knowledge. *Economic Geography*, *93*(1), 1–23.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99–120.

Bathelt, H., Malmberg, A., & Maskell, P. (2004). Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography*, *28*(1), 31–56.

Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of*

Sciences of the United States of America, *104*(17), 7301–6.

Boschma, R. (2005). Proximity and Innovation: A Critical Assessment. *Regional Studies, 39*(1), 61–74.

Brenner, T., Cantner, U. and H. Graf (2013). Introduction: Structure and dynamics of innovation networks. *Regional Studies* 47(5), 647-650.

Breschi, S., & Lenzi, C. (2016). Co-invention networks and inventive productivity in US cities. *Journal of Urban Economics*, *92*, 66–75.

Breschi, S., & Lissoni, F. (2009). Mobility of skilled workers and co-invention networks: An anatomy of localized knowledge flows. *Journal of Economic Geography*, *9*(4), 439–468.

Broekel, T., & Boschma, R. (2012). Knowledge networks in the Dutch aviation industry: the proximity paradox. *Journal of Economic Geography*, *12*(2), 409–433.

Burt, R. S. (1992). *Structural Holes*. Cambridge, MA: Harvard Business Press.

Burt, R. S. (2000). The network structure of social capital. *Research in Organizational Behavior*, *22*, 345–423.

Butts, C. T., Acton, R. M., Hipp, J. R., & Nagle, N. N. (2012). Geographical variability and network structure. *Social Networks*, *34*(1), 82–100.

Cantner, U. and H. Graf (2004). Cooperation and specialization in German technology regions. *Journal of Evolutionary Economics* 14(5): 543-562.

Cantner, U. and H. Graf (2011). Innovation networks: formation, performance and dynamics. In Anotonelli, C. (ed.) *Handbook on the Economic Complexity of Technological Change*. Cheltenham: Edward Elgar, pp 366-394.

Cantner, U., Meder, A., & Ter Wal, A. L. J. (2010). Innovator networks and regional knowledge base. *Technovation*, *30*(9), 496–507.

Capello, R. and C. Lenzi (2013). Territorial pattenrs of innovation: a taxonomy of innovative regions in Europe. *Annals of Regional Science* 51, 119-154.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly. Mar1990*, *35*(1), 128–152.

Cyert, R. M., & March, J. G. (1963). *A Behavioral Theory of the Firm*. Englewood Cliffs, NJ: Prentice Hall.

De Nooy, W., Mrvar, A., & Batagelj, V. (2011). *Exploratory Social Network Analysis with Pajek* (Vol. 27). Cambridge: Cambridge University Press.

Duranton, G., & Puga, D. (2001). Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review*, *91*(5), 1454–1477.

Engelsman, E. C., & van Raan, A. F. J. (1994). A patent-based cartography of technology.

Research Policy, *23*(1), 1–26.

Esposito, C. & Rigby, D. (2018). Buzz and pipelines: the costs and benefits of local and non-local interaction. *Journal of Economic Geography*, https://doi.org/10.1093/jeg/lby039.

Fleming, L., King, C., & Juda, A. (2007). Small Worlds and Regional Innovation. *Organization Science*, *18*(6), 938–954.

Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: Evidence from patent data. *Research Policy*, *30*(7), 1019–1039.

Gertler, M. S. (2003). Tacit knowledge and the economic geography of context, or The undefinable tacitness of being (there). *Journal of Economic Geography*, *3*(1), 75–99.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of Political Economy*, *100*(6), 1126–1152.

Goldberger, A.S. (1991). *A Course in Econometrics*. Cambridge, MA: Harvard University Press.

Gordon, I. & McCann, P. (2000). Industrial clusters: Complexes, agglomerations, and/or social networks? *Urban Studies* 37(3), 513-532.

Grabher, G. (1993). The weakness of strong ties: the lock-in of regional development in the Ruhr area. In, Grabher, G. (ed.) *The Embedded Firm: On the Socioeconomics of Industrial Networks*, 255–277.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology,* 78(6), 1360–1380.

Griliches, Z. (1990). Patent statistics as economic indicators: a survey. *Journal of Economic Literature*, *28*(December), 1661-1707.

Hagedoorn, J. (1993). Understanding the rationale of strategic technology partnering: Interorganizational modes of cooperation and sectoral differences. *Strategic Management Journal*, *14*(5), 371–385.

Healy, A., & Morgan, K. (2012). Spaces of innovation: Learning, proximity and the ecological turn. *Regional Studies*, *46*(8), 1041–1053.

Henderson, J. V. (2003). Marshall's scale economies. *Journal of Urban Economics*, *53*(1), 1–28.

Hirschman, A.O. (1964). The paternity of an index. *American Economic Review*, *54*(5), 761.

Jacobs, J. (1969). *The Economy of Cities*. New York: Random House.

Jaffe, A. B. (1986). Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value. *National Bureau of Economic Research - Working Paper*.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, *108*(3), 577–598.

Joo, S. H., & Kim, Y. (2010). Measuring relatedness between technological fields. *Scientometrics*, *83*(2), 435–454.

Kauffman, S. A. (1993). *The Origins of Order: Self Organization and Selection in Evolution*. Oxford: Oxford University Press.

Knoben, J., & Oerlemans, L. A. G. (2012). Configurations of inter-organizational knowledge links: Does spatial embeddedness still matter? *Regional Studies*, *46*(8), 1005–1021.

Kogler, D. F., Rigby, D. L., & Tucker, I. (2013). Mapping Knowledge Space and Technological Relatedness in US Cities. *European Planning Studies*, *21*(9), 1374–1391.

Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, *3*(3), 383–397.

Lai, R., D'Amour, A., Yu, A., Sun, Y., & Fleming, L. (2012). Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database (1975 - 2010). Harvard Dataverse. http://doi.org/hdl/1902.1/15705

Lobo, J., & Strumsky, D. (2008). Metropolitan patenting, inventor agglomeration and social networks: A tale of two effects. *Journal of Urban Economics*, *63*(3), 871–884.

Malmberg, A., & Maskell, P. (2006). Localized learning revisited. *Growth and Change*, *37*(1), 1–18.

Marshall, A. (1920). *Principles of Economics*. London: Macmillan.

Matula, D. W., & Beck, L. L. (1983). Smallest-last ordering and clustering and graph coloring algorithms. *Journal of the ACM*, *30*(3), 417–427.

Moodysson, J. (2008). Principles and practices of knowledge creation: On the organization of "buzz" and "pipelines" in life science communities. *Economic Geography* 84(4), 449-469.

Neffke, F., Henning, M., Boschma, R., Lundquist, K.-J., & Olander, L.-O. (2011). The Dynamics of Agglomeration Externalities along the Life Cycle of Industries. *Regional Studies*, *45*(1), 49–65.

Nesta, L., & Saviotti, P. P. (2005). Coherence of the knowledge base and the firm's innovative performance: evidence from the US pharmaceutical industry. *The Journal of Industrial Economics*, *53*(1), 123–142.

Penrose, E. (1959). *The Theory of the Growth of the Firm*. New York: John Wiley.

Potter, A., & Watts, H. D. (2011). Evolutionary agglomeration theory: increasing returns, diminishing returns, and the industry life cycle. *Journal of Economic Geography*, *11*(3), 417–455.

Powell, W. W. (1990). Neither market nor hierarchy. *Research in Organizational Behavior, 12*, 295–336.

Powell, W. W., Koput, K. W., & Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly*, 116–145.

Rigby, D. L., & Brown, W. M. (2015). Who benefits from agglomeration? *Regional Studies*, *49*(1), 28–43.

Rowley, T., Behrens, D., & Krackhardt, D. (2000). Redundant governance structures: An analysis of structural and relational embeddedness in the steel and semiconductor industries. *Strategic Management Journal*, 369–386.

Samila, S. & Sorenson, O. (2011). Venture capital, entrepreneurship, and economic growth. *The Review of Economics and Statistics* 93(1), 338-349.

Saxenian, A. (1994). *Regional advantage: culture and competition in Silicon Valley and route 128*. Cambridge, MA: Harvard University Press.

Seidman, S. B. (1983). Network Structure and Minimum Degree. *Social Networks*, *5*, 269–287.

Singh, J. (2005). Collaborative Networks as Determinants of Knowledge Diffusion Patterns. *Management Science*, *51*(5), 756–770.

Sorenson, O. (2005). Social networks and industrial geography. In *Entrepreneurships, the New Economy and Public Policy* (pp. 55–69). Springer.

Strumsky, D., & Thill, J.-C. (2013). Profiling Us Metropolitan Regions By Their Social Research Networks and Regional Economic Performance. *Journal of Regional Science*, *53*(5), 813–833.

Ter Wal, A. (2013). Cluster emergence and network evolution: A longitudinal analysis of the inventor network in Sophia-Antipolis. *Regional Studies* 47(5), 651-668.

Trippl, M., Todtling, F. and L. Lengauer (2009). Knowledge sourcing beyond buzz and pipelines: evidence from the Vienna software sector. Economic Geography 85(4), 443-462.

Uzzi, B. (1996). The sources and consequences of embeddedness for the economic performance of organizations: The network effect. *American Sociological Review*, *61*(4), 674–698.

Uzzi, B. (1997). Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness. *Administrative Science Quarterly*, *42*(1), 35–67.

Van der Wouden, F. (2018), Co-inventors on historical US patents: Changing Patterns in Collaboration, Complexity and Geography, *UCLA Geography of Innovation Working Paper*

Walker, G., Kogut, B., & Shan, W. (1997). Social capital, structural holes and the formation of an industry network. *Organization Science*, *8*(2), 109–125.

Wernerfelt, B. (1984). A Resource based view of the firm. *Strategic Management Journal*, *5*(2), 171–180.

Whittington, K. B., Owen-Smith, J., & Powell, W. W. (2009). Networks, Propinquity, and Innovation in Knowledge-intensive Industries. *Administrative Science Quarterly*, *54*(1), 90–

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–9.